JNCI Journal of the National Cancer Institute

Journal of the National Cancer Institute, Vol. 95, No. 6, E2, March 19, 2003

ONLINE PERSPECTIVE

Epidemiologic Studies of a Necessary Causal Risk Factor: Human Papillomavirus Infection and Cervical Neoplasia

Mark H. Schiffman, Philip Castle

Affiliation of authors: M. H. Schiffman, P. Castle, Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute (NCI), Bethesda, MD.

Correspondence to: Mark H. Schiffman, M.D., M.P.H., National Institutes of Health, Executive Plaza South, Rm. 7066, 6120 Executive Blvd., Rockville, MD 20892 (e-mail: schiffmm@epndce.nci.nih.gov).

In the 1990s, interdisciplinary research teams showed that human papillomavirus (HPV) infection causes virtually all cases of cervical cancer and its preinvasive cytopathologic precursors. Two such studies from the Journal are discussed here. Bosch et al. *(1)* demonstrated that virtually all cases of invasive cervical cancer worldwide contain DNA of the same 13 oncogenic HPV types. Schiffman et al. *(2)* showed that oncogenic HPV infection is so tightly linked to the diagnosis of preinvasive cytopathologic precursors that statistical adjustment for HPV infection explains previously established sexual risk factors for cervical neoplasia. It was important to confirm that HPV is the necessary cause of cervical cancer, which annually kills 200,000 women worldwide. No similarly important cancers have such clearly known necessary causes. Few other carcinogens are as powerful as these small, deceptively simple viruses. Many years later, prevention researchers are already introducing clinically useful HPV DNA tests *(3)*, and vaccination research is well under way *(4)*.

This short perspective touches on a few of the unique epidemiologic aspects of establishing HPV as a necessary causal agent and the impact of that discovery on subsequent work. The perspective is

personal, and the advances described below belong to a sizable community of researchers whom we cannot adequately acknowledge here.

Bosch et al. *(1)* found HPV DNA in nearly all cervical cancers because of two critical, under-appreciated methodologic triumphs. First, the huge collaborative epidemiologic team was able to collect more than 1000 good-quality frozen tumor specimens from women in 22 countries. Routine formalin-fixed tissues would not have permitted sensitive testing. A smaller effort restricted to the usual research centers would have left questions regarding global generalizability. The logistical effort, for which the International Agency for Research on Cancer (IARC) deserves credit, immediately established the universal conclusion that the same HPV types are associated with cancer everywhere.

Second, the HPV laboratory collaborators had painstakingly optimized the sensitive polymerase chain reaction (PCR) detection of a broad range of HPV types. The most critical aspects of epidemiology are proper classification of exposure and disease and proper choice of comparison groups. Laboratory scientists had originally suggested around 1980 that HPV types 16 and 18 cause cervical cancer *(5)*, based on epidemiologic and pathology evidence for a sexually transmitted etiologic agent. But at the birth of the field of HPV and cervical cancer, only a few HPV types had been characterized. PCR had not even been invented. The subsequent optimization of consensus PCR primers for HPV DNA took years from first concept to acceptable performance *(6)*. These methodologic efforts made possible the study by Bosch et al. *(1)* because they reduced both false-negatives (from limited type range or low analytic sensitivity) and false-positives (due to the amplification product contamination that had plagued some early PCR projects).

The percentage of cervical cancers found to contain oncogenic HPV DNA is, theoretically, the upper bound on the percentage caused by HPV. Even if all cervical cancers had demonstrable HPV DNA, however, the cervical cells of control women might also be HPV positive, making HPV just a ubiquitous benign virus. Some early PCR studies suggested the possibility of benign HPV ubiquity, and the concept of *etiologic fraction* formalizes this concern *(7)*. The etiologic fraction is the percentage of HPV-positive case patients adjusted downward to account for the HPV positivity in control subjects. Studying control women, with the necessary adaptation of HPV testing to noninvasive cervical cell collections, was much more difficult than studying women with cervical cancer. But once testing of control women was achieved, population surveys revealed that only a small minority of women in the general population are HPV infected at any one time. Therefore, the correction of the etiologic fraction estimated by the case series of Bosch et al. is negligible. In fact, HPV infection is now estimated to cause 99% of cervical cancers *(8)*.

While Bosch et al. *(1)* focused on cervical cancers, Schiffman et al. *(2)* concentrated on precursors of cervical cancer. The two efforts were designed to be complementary. Schiffman et al. studied cervical intraepithelial neoplasia as a surrogate endpoint for cervical cancer. The pathogenesis of cervical cancer was already known to develop from reasonably well-defined pathologic precursors, identifiable by histology or cytology (Pap tests).

Specifically, Schiffman et al. *(2)* were able to show that HPV infection explains the risk of cervical intraepithelial neoplasia associated with previously established sexual risk factors for cervical cancer, particularly lifetime number of sexual partners (the epidemiologic variable that first suggested a sexually transmitted causal agent). But HPV infection does not explain causal cofactors, like smoking and parity, and we have continued to study these factors.

The epidemiologic study of cervical intraepithelial neoplasia and HPV has demonstrated the importance of proper exposure measurements and the critical role of refining difficult pathologic distinctions. The more clearly we have defined cervical precancerous conditions by expert review or refined the HPV test data by use of successively more sensitive and specific assays, the stronger the estimates of HPV oncogenicity have become *(9)*.

To confirm that the amazingly strong association between oncogenic HPV and cervical neoplasia seen by Bosch et al. *(1)* and Schiffman et al. *(2)* was prospective and did not result from "reverse causality" (i.e., cancer leading to increased carriage or detection of HPV), a series of long-term prospective studies have been completed, including a 20,000-woman cohort in Oregon *(10)*, a 10,000-woman cohort in Costa Rica *(11)*, and similar large projects in England *(12)*, Brazil *(13)*, Denmark *(14)*, and California *(15)*. The prospective data *(16,17)* and archival cohort studies *(18)* corroborate the findings of Bosch et al. completely, in that HPV 16 and the other oncogenic types defined by the IARC case series have been confirmed prospectively to be causal, while other HPV types have been shown to be benign.

It is now known that cervical cancer arises via three carcinogenic steps. *HPV infection* of the cervix occurs mainly as a result of sexual intercourse. The resultant "phenotype" of the HPV-infected cervical cell is highly variable (e.g., cytologic diagnoses of atypical squamous cells of unknown significance [ASCUS] and low-grade intraepithelial lesions [LSIL]). Most infections tend to resolve over a 1- to 2-year period. A small percentage, however, undergo *progression* to a precancerous condition, with high risk of subsequent *invasion.*

The implications of accepting oncogenic HPV as the necessary but not sufficient cause of cervical cancer are worth discussing. Epidemiology is designed to study risk factors, and the heart of epidemiology is the concept of *relative risk.* Epidemiologists wished to know how much more likely oncogenic HPV-infected women were to develop cervical cancer than noninfected women. The relative risk is computed from a standard two-by-two table as the absolute risk of cervical cancer following HPV infection divided by the absolute risk in the absence of HPV infection. When the absolute risk among uninfected women approaches zero, however, the relative risk asymptotically "balloons" to infinity and the standard statistical tables must be reconsidered.

For example, when studying other possible cervical cancer risk factors, such as smoking, we cannot use usual approaches to adjust for confounding (e.g., Mantel–Haenszel or logistic regression) and effect modification, because there are almost no truly uninfected case patients. Almost all of the seemingly uninfected case patients are misclassified, and they should, therefore, be excluded from averaged risk

estimates. Because HPV is a necessary cause of cervical cancer, selection of the proper controls in case–control studies is also profoundly affected. Control subjects should properly represent the population at risk of cervical cancer, and women without exposure to oncogenic HPV types are not at risk of cervical cancer. Consequently, conventional population-based approaches may not be the best ways for studying cervical cancer risk factors that might influence HPV-induced carcinogenesis.

Specifically, most epidemiologic work is now focused on progression. Because HPV infection is extremely common over a lifetime, epidemiologists are now trying to understand why a small percentage of infected women have persistent and progressive disease, rather than the much more common pattern of regression. A number of potential HPV cofactors have been suggested, including cell-mediated immunity, genetic factors, smoking, parity, oral contraceptive use, cervical inflammation, and *Chlamydia trachomatis* infection *(19–21)*. All, with perhaps the exceptions of smoking and parity, have been plagued by weak and/or inconsistent findings. Methodologic issues may be partly to blame.

In the study of possible cofactors, proper adjustment for oncogenic HPV infection is critical. Any factor that is even slightly associated with risk of oncogenic HPV infection will appear to influence the risk of cervical neoplasia. This powerful confounding must be addressed. Two approaches have been used: 1) multivariate modeling (by stratification or regression models) and 2) restricting the analysis to only those women infected with oncogenic HPV types (HPV-restricted analysis). Both methods have limitations. In the case of multivariate modeling, if no case patients are truly HPV negative, then modeling will incorporate artifacts as the result of misclassification of HPV status. HPV-restricted analysis is perhaps the only certain means to completely control for the effects of HPV. Certainly, these women represent the true at-risk group. However, for this approach to be useful it is necessary to understand more precisely what it means to find HPV in control women, particularly in cross-sectional studies. Women with a precancerous condition and especially those with cancer are older, on average, than women with incident HPV infection. Therefore, it is important to control for age. However, HPV DNA-positive control subjects who are age-matched to case patients may be older women (compared with all infected women) who have recently acquired the HPV infection as a result of the same sexual behaviors linked to the acquisition of other sexually transmitted diseases (STDs) or smoking, leading to a biased estimate of STDs and smoking as HPV cofactors. Or, these women might have a persistent infection and are, therefore, already at elevated risk of disease progression, possibly biasing any association of an HPV cofactor toward the null.

One methodologic approach to overcome these limitations is to add an element of time to the analysis. This can be accomplished in a variety of ways, none of which has been shown to be clearly superior to the others. One method is to use younger, HPV-positive control subjects as the comparison group to the case patients. The disease in most of these women will regress rather than progress, and in this way the case patients are being compared with women who are at an earlier stage in the natural history of cervical cancer. This approach assumes, perhaps erroneously, that that there is no trend over time in the characteristics of HPV-infected younger women. Another method is to use serology as a measure of past HPV exposure to simulate acquisition and clearance of infection *(22)*, although serology is insensitive, detecting only about one half of current infections. For example, a possible control group

might include women who are seropositive (for an oncogenic HPV type), HPV DNA negative, and cytologically normal; these women have had, and then cleared, an infection. Alternatively, cytologically normal women with a history of untreated LSIL might also be used as control subjects because LSIL is a manifestation of HPV infection; however, there is no way to ascertain whether or not the infection was due to an oncogenic HPV type.

It seems intuitively obvious that prospective studies could optimally assess causal associations between cofactors and HPV progression. But conventional prospective studies also present some of the same concerns found in case–control studies, namely uncertainty regarding the timing of infection ("left censoring"). Again, surrogates of past infection, such as serology, might be considered to overcome this problem.

In current studies, several epidemiologic groups are emphasizing prospective studies, but those with repeated measurements of large cohorts of women. Our group measures HPV infection and stages in cervical carcinogenesis using visual, cytologic, molecular, and immunologic methods. In attempting to identify and validate new biomarkers of risk, we are no longer discussing the simple risk of HPV and cancer. Instead, cervical carcinogenesis is divided into its steps (infection, progression [versus regression], and invasion), and each transition is analyzed separately. For example, women with precancerous conditions are compared with those with acute infection to examine risk factors for disease progression. It should be noted that uninfected women would not be good controls for these comparisons. Studies of repeated measurements of risk biomarkers and intermediate histopathologic endpoints require sophisticated and difficult statistical tests, some of which are incompletely developed. Collaborations with statisticians are increasingly important to derive maximum information from the cohorts.

In conclusion, epidemiologic studies have revealed not only that women without HPV do not get cervical cancer but also that most women with HPV do not get cervical cancer. A next generation of biomarkers is needed that will distinguish infections with oncogenic types that are truly predictive of risk from those that are not. Again, such discoveries will require a coordinated effort between laboratory scientists and epidemiologists. It is easy to state this new epidemiologic goal, but admittedly it is a bit like "leaving home" to step beyond the stratospheric relative risks of greater than 100 associating HPV with cervical cancer as we face the extreme complexity of the next series of intensive, molecular epidemiologic studies.

REFERENCES

(1) Bosch FX, Manos MM, Munoz N, Sherman M, Jansen AM, Peto J, et al. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International Biological Study on Cervical Cancer (IBSCC) Study Group. J Natl Cancer Inst 1995;87:796–802.[Abstract]

(2) Schiffman MH, Bauer HM, Hoover RN, Glass AG, Cadell DM, Rush BB, et al. Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. J Natl Cancer Inst 1993;85:958–64.[Abstract]

(3) Solomon D, Schiffman M, Tarone R. Comparison of three management strategies for patients with atypical squamous cells of undetermined significance: baseline results from a randomized trial. J Natl Cancer Inst 2001;93:293–9.[Abstract/Free Full Text]

(4) Harro CD, Pang YY, Roden RB, Hildesheim A, Wang Z, Reynolds MJ, et al. Safety and immunogenicity trial in adult volunteers of a human papillomavirus 16 L1 virus-like particle vaccine. J Natl Cancer Inst. 2001;93:284–92.[Abstract/Free Full Text]

(5) Durst M, Gissmann L, Ikenberg H, zur Hausen H. A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions. Proc Natl Acad Sci USA 1983;80:3812–5.[Medline]

(6) Hildesheim A, Schiffman MH, Gravitt PE, Glass AG, Greer CE, Zhang T, et al. Persistence of type-specific human papillomavirus infection among cytologically normal women. J Infect Dis 1994;169:235–40.[Medline]

(7) Schatzkin A, Freedman LS, Schiffman MH, Dawsey SM. Validation of intermediate end points in cancer research. J Natl Cancer Inst 1990;82:1746–52.[Abstract]

(8) Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol 1999;189:12–9. [CrossRef][Medline]

(9) Schiffman MH, Schatzkin A. Test reliability is critically important to molecular epidemiology: an example from studies of human papillomavirus infection and cervical neoplasia. Cancer Res 1994;54(7 Suppl):1944s–1947s.[Abstract]

(10) Liaw KL, Glass AG, Manos MM, Greer CE, Scott DR, Sherman M, et al. Detection of human papillomavirus DNA in cytologically normal women and subsequent cervical squamous intraepithelial lesions. J Natl Cancer Inst 1999;91:954–60.[Abstract/Free Full Text]

(11) Herrero R, Hildesheim A, Bratti C, Sherman ME, Hutchinson M, Morales J, et al. Population-based study of human papillomavirus infection and cervical neoplasia in rural Costa Rica. J Natl Cancer Inst 2000;92:464–74.[Abstract/Free Full Text]

(12) Deacon JM, Evans CD, Yule R, Desai M, Binns W, Taylor C, et al. Sexual behaviour and smoking as determinants of cervical HPV infection and of CIN3 among those infected: a case–control study nested within the Manchester cohort. Br J Cancer 2000;83:1565–72.[CrossRef][Medline]

(13) Franco EL, Villa LL, Sobrinho JP, Prado JM, Rousseau MC, Desy M, et al. Epidemiology of acquisition and clearance of cervical human papillomavirus infection in women from a high-risk area for cervical cancer. J Infect Dis 1999;180:1415–23.[CrossRef][Medline]

(14) Kjaer SK, van den Brule AJ, Bock JE, Poll PA, Engholm G, Sherman ME, et al. Human papillomavirus—the most significant risk determinant of cervical intraepithelial neoplasia. Int J Cancer 1996;65:601–6.[CrossRef][Medline]

(15) Moscicki AB, Hills N, Shiboski S, Powell K, Jay N, Hanson E, et al. Risks for incident human papillomavirus infection and low-grade squamous intraepithelial lesion development in young females. JAMA 2001;285:2995–3002.[Abstract/Free Full Text]

(16) Ho GY, Bierman R, Beardsley L, Chang CJ, Burk RD. Natural history of cervicovaginal papillomavirus infection in young women. N Engl J Med 1998;338:423–8.[Abstract/Free Full Text]

(17) Nobbenhuis MA, Walboomers JM, Helmerhorst TJ, Rozendaal L, Remmink AJ, Risse EK, et al. Relation of human papillomavirus status to cervical lesions and consequences for cervical-cancer screening: a prospective study. Lancet 1999;354:20–5.[CrossRef][Medline]

(18) Wallin KL, Wiklund F, Angstrom T, Bergman F, Stendahl U, Wadell G, et al. Type-specific persistence of human papillomavirus DNA before the development of invasive cervical cancer. N Engl J Med 1999;341:1633–8.[Abstract/Free Full Text]

(19) Schiffman MH, Brinton LA. The epidemiology of cervical carcinogenesis. Cancer 1995;76 (10 Suppl):1888–901.[Medline]

(20) Hildesheim A, Herrero R, Castle PE, Wacholder S, Bratti MC, Sherman ME, et al. HPV co-factors related to the development of cervical cancer: results from a population-based study in Costa Rica. Br J Cancer 2001;84:1219–26.[CrossRef][Medline]

(21) Castle PE, Hillier SL, Rabe LK, Hildesheim A, Herrero R, Bratti MC, et al. An association of cervical inflammation with high-grade cervical neoplasia in women infected with oncogenic human papillomavirus (hpv). Cancer Epidemiol Biomarkers Prev 2001;10:1021–7.[Abstract/Free Full Text]

(22) Anttila T, Saikku P, Koskela P, Bloigu A, Dillner J, Ikaheimo I, et al. Serotypes of Chlamydia trachomatis and risk for development of cervical squamous cell carcinoma. JAMA 2001;285:47–51. [Abstract/Free Full Text]